Biological Journal of the Linnean Society, 2007, 91, 523–539. With 5 figures

Species-richness of the *Anopheles annulipes* complex (Diptera: Culicidae) revealed by tree and model-based allozyme clustering analyses

DESMOND H. FOLEY[1,2]*, JOAN H. BRYAN[2] and RICHARD C. WILKERSON[1]

[1]*Department of Entomology, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910, USA*
[2]*Tropical Health Program, Australian Centre for International and Tropical Health and Nutrition, The University of Queensland, Brisbane, Queensland 4072, Australia*

Received 24 January 2006; accepted for publication 15 September 2006

The Australasian *Anopheles annulipes* complex contains at least ten sibling species, some of which are important vectors of myxomatosis in rabbits. We aimed to establish how many species occurred among specimens from 61 sites throughout Australia, scored for 32 putative allozyme loci. We compared the number of species predicted from tree-based clustering of operational taxonomic units (OTUs) with that from a novel model-based Bayesian clustering approach for individual genotypes. We rejected the hypothesis of conspecificity of OTUs if they differed by at least 20% fixed differences and 0.300 Nei's standard genetic distance *D*. According to these criteria, 18–25 species occur, making this the most species-rich anopheline complex known to date. A conservative estimate from the Bayesian analysis was 15–20 species. There was large overlap in the assignment of individuals to clusters inferred from the Bayesian and tree-based analyses. The genetic clustering of northern and southern distributed species and an apparent cline in alleles of the enzyme glucose phosphate isomerase suggest that a latitude-dependent factor, such as temperature, may have played a role in speciation and the subsequent distribution of species. Ecological niche modelling of clusters predicted that none occur in New Guinea, emphasizing that additional, as yet unsampled, species may occur.

ADDITIONAL KEYWORDS: Australia – Bayesian – cline – ecological niche modelling – mosquito – myxomatosis – sibling species – species radiation.

INTRODUCTION

Anopheles annulipes s.l. Walker (subgenus *Cellia*, Neomyzomyia series) is the most ubiquitous anopheline in Australia and also occurs in New Guinea (Lee *et al.*, 1987). *Anopheles annulipes s.l.* has been implicated in past malaria outbreaks in Australia (Black, 1972), is the most important vector of myxomatosis in many areas of Australia (Fenner & Ratcliffe, 1965; Parer & Korn, 1989), and a number of other arboviruses have been recovered from this taxon (Russell, 1995). This taxon exhibits extensive morphological variation that has resulted in various taxonomic interpretations; five names have been synonymized under *An. annulipes* (*Anopheles musivus* Skuse, *Anopheles mastersi* Skuse, *Anopheles perplexus* Taylor, *An. perplexus* var. *persimilis* Taylor, and *Anopheles derricki* Taylor). Cross-mating evidence and polytene chromosomal typing suggest that *An. annulipes s.l.* is composed of at least ten sibling species, seven of which were given the letter designations A to G (Booth & Bryan, 1986). At least two of these chromosomal types (sp. A and sp. G) do not interbreed in nature (confirming their status as biological species) and have different ecologies (Bryan *et al.*, 1991; Foley & Bryan, 1991a, b; Foley, Barnes & Bryan, 1992). A phylogeny of Australasian anophelines, including four species of the *An. annulipes* complex, based on sequences of the cytochrome oxidase subunit *II* gene, suggests that the *An. annulipes* complex is

*Corresponding author. E-mail: foleydes@si.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2007		**00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Species-richness of the Anopheles annulipes complex (Diptera: Culicidae) revealed by tree and model-based allozyme clustering analyses	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Army Institute of Research,Department of Entomology,Silver Spring,MD,20910**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**17**	

monophyletic (Foley *et al.*, 1998). The delineation of species boundaries is fundamental to any further study of this taxon, and past studies will have to be reassessed in the light of its multispecies status.

Detection of mixtures of species of sexually reproducing organisms in sympatry is possible through the observation of fixed differences in codominant genotypes, which indicate assortative mating. According to the phenetic approach, a predetermined value for inter- and intraspecific genetic distance is applied to genetic distances between operational taxonomic units (OTUs) from allopatric sites to determine the number of putative species. Various clustering approaches for OTUs are available, including tree-based methods such as the unweighted pair group method of analysis (UPGMA) and Neighbour-joining (NJ) (Saitou & Nei, 1987).

By contrast, a Bayesian approach, as implemented in the program STRUCTURE 2.0 (Pritchard, Stephens & Donnelly, 2000), analyses individual genotypes, by estimating the likelihood of an individual's membership among each of a predefined number of clusters (K). Ideally, estimates of the posterior probability, ln(K) (Evanno, Ragnaut & Coudet, 2005), plateaus with increasing K once the real number of groups is reached (Pritchard & Wen, 2003). When this occurs, the K that matches the real number of groups is often the lowest of the likelihood scores in the plateau (Pritchard & Wen, 2003). STRUCTURE normally has been applied to questions of intraspecific structure (Rosenberg *et al.*, 2001) but D. H. Foley (unpubl. data) showed that this approach successfully identified the correct number of species among a group of simulated genotype data, and actual allozyme data for ten species of the *An. punctulatus* complex (Foley, Cooper & Bryan, 1995). This Bayesian approach shows promise as a new method for determining the number of species among a group of genotypes.

We collected allozyme data for *An. annulipes s.l.* from locations around Australia to estimate the number of species using the model-based Bayesian approach and the tree-based UPGMA and NJ clustering methods. The results obtained suggest that far more species occur in the *An. annulipes* complex than was previously suspected.

MATERIAL AND METHODS

MOSQUITO COLLECTIONS AND IDENTIFICATION

Mosquitoes were collected between the early 1980s to the mid-1990s, either as larvae that were reared to adults or as adults from CO_2-baited light traps or from human landing catches (Table 1). Adult females were identified as *An. annulipes s.l.* using the morphological keys of Lee *et al.* (1987), and specimens were stored at −80 °C. Specimens identified on their chromosomes as *An. annulipes* sp. A and sp. G were included from Mildura, Victoria (VIC) and Griffith, New South Wales (NSW), respectively. Specimens from Homebush, Termeil State Forest, and Lord Howe Island, NSW were identified on chromosomes as sp. C. Chromosomally-typed specimens from Mataranka, Northern Territory (NT), from the study of Booth & Bryan (1986), were also included.

ALLOZYME ELECTROPHORESIS

Cellulose acetate (CA) allozyme electrophoresis was carried out as described previously (Foley *et al.*, 1993; Foley, Meek & Bryan, 1994). Specimens of *Anopheles farauti* Laveran (= *An. farauti* No. 1), *Anopheles hinesorum* Schmidt (= *An. farauti* No. 2), and *Anopheles torresiensis* Schmidt (= *An. farauti* No. 3) from colonies that were maintained at the Army Malaria Research Unit, Ingleburn, NSW, Australia were used as controls for band migration distance. Loci of the lowest anodic mobility in a zymogram were numbered '1' and the slowest allelomorphs were designated 'a'.

The 24 allozymes used in this study were aconitate hydratase (ACON, EC no. 4.2.1.3), acid phosphatase (ACP, EC no. 3.1.3.2), adenylate kinase (AK, EC no. 2.7.4.3), α-amylase (αAMY, EC no. 3.2.1.1), enolase (ENOL, EC no. 4.2.1.17), fructose-1,6-diphosphatase (FDP, EC no. 3.1.3.11), glutamate-oxaloacetate transaminase (GOT, EC no. 2.6.1.1), glucose-phosphate isomerase (GPI, EC no. 5.3.1.9), β-hydroxybutyrate dehydrogenase (HBDH, EC no. 1.1.1.30), β-galactosidase (βGAL, EC no. 3.2.1.23), α-glycerophosphate dehydrogenase (αGPD, EC no. 1.1.1.8), hexokinase (HK, EC no. 2.7.1.1), isocitric dehydrogenase (IDH, EC no. 1.1.1.42), lactate dehydrogenase (LDH, EC no. 1.1.1.27), malate dehydrogenase (MDH, EC no. 1.1.1.37), malic enzyme (ME, EC no. 1.1.1.40), mannose-6-phosphate isomerase (MPI, EC no. 5.3.1.8), octanol dehydrogenase (ODH, EC no. 1.1.1.73), peptidase B (PEPB, EC no. 3.4.13.9), peptidase D (PEPD, EC no. 3.4.13.9), phosphoglucomutase (PGM, EC no. 2.7.5.1), 6-phosphogluconate (6PGD, EC no. 1.1.1.44), pyruvate kinase (PK, EC no. 2.7.1.40) and L-threonine 3-dehydrogenase (THDH, EC no. 1.1.1.103). Hexokinase exhibited three zones of activity and it was assumed that each band was controlled by a separate locus.

TREE-BASED ANALYSIS

Allozyme data from each collection site were inspected for groups of individuals that differed by more than one fixed allelic difference. This pattern indicates assortative mating and the presence of two or more OTUs, especially when multiple samples of each

Table 1. Collection details for an allozyme study of 366 *Anopheles annulipes s.l.* in Australia

Number	Locality name*	Longitude	Latitude	OTU (*N*)	15% Cluster	*K*15¶	*K*20	*K*25
1	Basalt R.†	145°46′E	19°37′S	33(2),34,35(2),36	6,17	4,9	5,12	1,19
2	Cairns region (site A)†	145°34′E	16°53′S	1(2),2,3(5)	1,7,20	7,8,?	1,2,?	3,4,20,?
3	Cairns region (site B)†	145°33′E	16°58′S	4,5(2)	19,20	8	2	20
4	Cairns region (site C)†	145°29′E	17°18′S	6	7	1	7	6
5	Clarke R.†	145°26′E	19°13′S	42(3),43(3),44,45(2),46	6,8,10,17	4,9,12,14	5,12,16,17	1,7,19,24
6	Clermont†	147°38′E	22°50′S	47(2),48(2),49(4)	10,17	4,12	12,16	19,24
7	Dilulu†	150°16′E	23°53′S	28(3),29	17	4	12	19
8	Emerald‡	148°10′E	23°31′S	30(2),31,32	5,10,17	3,4,12	12,16,?	14,19,24
9	Eidsvold†	151°07′E	25°22′S	23(3),24,25,26,27	3,5,11	3,5,15	6,8,14	18,21,22,?
10	Eungella‡	148°30′E	21°08′S	12(4)	1	7	1	16
11	Horn Island†‡	142°17′E	10°35′S	14(2),15(4),16(5),17(3)	18	8	2,20,?	2,15,?
12	Innot Hot Springs†	145°14′E	17°40′S	38(2),39(2),40,41	8,10,17	4,12,14	12,16,17	7,19,24
13	Kennedy Creek‡	144°26′E	15°43′S	7(2)	20	12	16	24
14	Lake Manchester†	152°45′E	27°28′S	20(5),21,22(2)	1,4	7,15	1,8	3,18
15	Point Stewart Road‡	143°41′E	14°04′S	11	20	8	2	4
16	Prince of Wales Island†	142°07′E	10°43′S	13(10)	18	8	20	2
17	Ravenshoe‡	145°29′E	17°36′S	37	2	7	1	3
18	Silver Plains‡	143°33′E	13°59′S	18(2),19(2)	12,20	8,12	2,16	4,25
19	Townsville region†‡	146°45′E	19°26′S	50(2),51(4),52,53(5), 54(2),55(8),56(5)	9,13,14, 15,16	10,14	9,17	7,12,23
20	Yungaburra†	145°35′E	17°16′S	8(6),9,10	12	8	2	4
21	Bateman's Bay‡	150°15′E	35°44′S	77,78,79	21,28	13,15	13,?	13,?
22	Condobolin‡	147°09′E	33°05′S	69(3)	26	3	14	22
23	Conjola‡	150°26′E	35°13′S	57(4),58	28	13	11	9
24	Dunoon†	153°19′E	28°41′S	60,61(2),62,80, 81(3),82	21,24,30, 33,34	7,15,?	1,8,13,?	3,13,16,18
25	Forbes‡	148°01′E	33°23′S	73(5)	26	3	14	22
26	Grafton (26 km south east)†	152°58′E	29°50′S	66(2)	25	7	1	3
27	Griffith‡	146°02′E	34°17′S	86	22	2	19	8
28	Hanwood‡	146°02′E	34°20′S	67(5)	22	2	19	8
29	Homebush	151°05′E	33°52′S	63(4)	29	13	11	9
30	Lake Cargelligo‡	146°22′E	33°18′S	68(5)	26	3	14	22
31	Lord Howe Island†	159°05′E	31°33′S	76(4)	29	13	11	9
32	McCarrs Creek†	151°16′E	33°40′S	83(8)	34	6,15	8,18	10,18
33	Menidee‡	142°25′E	32°24′S	72(5)	26	3	14	22
34	Mittagong†	150°27′E	34°27′S	75(5)	34	15	8	18
35	Tenterfield (Reedy Creek)†	151°50′E	29°09′S	64,65(3)	23,33	9,15	5,8	1,18
36	Termeil State Forest‡	150°22′E	35°26′S	84(5)	28	13	11	9
37	Walgett‡	148°07′E	30°01′S	70(2),71	22,26	2,3	14,19	8,22
38	Warren‡	147°50′E	31°42′S	74(3)	26	3	14	22
39	Woronora†	151°02′E	34°02′S	59(6)	33	15	8	18

Table 1. *Continued*

Number	Locality name*	Longitude	Latitude	OTU (*N*)	15% Cluster	*K*15¶	*K*20	*K*25
40	Billiluna (91K51I5-6, 91K52I3)‡	127°40′E	19°33′S	127(3),128(2),129, 130,135(5)	41,42,45	1,5,?	6,7,?	4,6,21,25,?
41	Broome (91 WEEK39)‡	122°14′E	17°58′S	106 107	40,41	1,?	7,?	6,21
42	Busselton (BSN37–39)‡	115°21′E	33°39′S	118(2),119(2),120, 131,132(5)	35,36, 38,41	1,3,?	3,7,10,?	6,15,17,?
43	Capel Shire‡ (BSN49,60–64,67,80,81)	115°33′E	33°33′S	109(3),110(3),111(3), 112(2),113,114,115, 116,117,121(2),122, 133(2),134(3),138(5)	35,36, 37,39	3	3,10,14,?	15,17,22,?
44	Kalumburu (91K11)‡	126°38′E	14°18′S	104	47	8	2	?
45	Kununurra (10544–5)‡	128°44′E	15°46′S	108(3),136(3),137(5)	41	1	7	6
46	Minnie R. Derby Shire‡ (91 week 23–25)	123°36′E	17°47′S	105(3),123,124(2), 125 126 44,45	42,43,	5,?	6,20	2,21
47	Alice Springs (Il Parpa Swamp)‡	134°26′E	24°10′S	88(6),89(5)	35,47	3,8,?	9,15,20,?	11,14
48	Berry Springs‡	130°58′E	12°42′S	90(3),91(2)	46,47	8,11	2,4,?	5,11,20,?
49	Darwin‡§	130°50′E	12°27′S	94,95(2),96,97,98(2), 99,100,101,102,103	41,46,47	1,8,11	2,4,7	5,6,20,?
50	Jim Jim Creek§	133°05′E	13°05′S	92(2),93(6)	46,47	8,11	2,4,?	5,20,?
51	Mataranka‡	133°04′E	14°56′S	87(3)	42	12	16	24
52	Avon R. Shire (Woodpile)‡	147°23′E	38°03′S	139(4)	28	13	11	9
53	Gunbower‡	144°22′E	35°58′S	143	26	3	14	22
54	Holland landing‡	147°28′E	38°04′S	141(3)	28	13	11	9
55	Marley Point‡	147°15′E	38°05′S	140(3)	28	13	11	9
56	Meerlieu‡	147°23′E	38°01′S	142(2)	28	13	11	9
57	Mildura†‡	142°10′E	34°11′S	85(4)	26	3	10,14,15	14,15
58	Bagdad R.‡	147°17′E	42°42′S	149 150	34	6	18	10
59	Devils Creek†	148°15′E	41°30′S	146,147(4),148	34	6	18	10
60	Glencoe Swamp†	148°15′E	41°30′S	144,145(2)	31,34	6	18,?	10,?
61	Upper Turners Marsh†	147°13′E	41°26′S	151,152(3)	27,32	6,?	18,?	10,?

Operational taxonomic units (OTUs) are given with sample size (if greater than one) as well as geographical distribution of 47 clusters comprising OTUs that differed by < 15%FD (for details, see Appendix). Assignment of individuals to *K* groups is given according to separate Bayesian analyses at $K = 15$, 20 and 25.

*Locations by State and Territory are: 1–20 (Queensland), 21–39 (New South Wales), 40–46 (Western Australia), 47–51 (Northern Territory), 52–57 (Victoria), 58–61 (Tasmania).

†Specimens collected as larvae.

‡Specimens collected by EVS trap (BioQuip).

§Specimens collected by night landing catches.

¶'?' indicates presence of at least one unassigned individual.

OTU occur sympatrically (Richardson, Baverstock & Adams, 1986). If no evidence for assortative mating was obtained then all specimens from a site were regarded as belonging to the same OTU, unless they differed by greater than three fixed differences, in which case specimens were treated conservatively, as separate OTUs. The percentage of loci for which OTUs do not share alleles [the percentage of fixed differences (%FD) between OTUs] and Nei's standard genetic distance D, corrected for small sample size (Nei, 1978), were calculated using the program BIGMAT (M. Adams, unpubl. data). For samples that were not scored for all enzymes, Nei's D and %FD were calculated from the data available. In addition to an Australia-wide analysis, regional analyses were performed for Queensland (QLD), NSW plus Tasmania (TAS) plus VIC and Western Australia (WA) plus NT.

Operational taxonomic units were regarded as separate species when they differed by at least 20%FD and/or 0.300 Nei's D. These levels of genetic divergence are based on those for the Australasian *An. punctulatus* complex in which the upper limits of intraspecific variation was 18%FD and 0.368 Nei's D (Foley *et al.*, 1993). However, this level was inflated by one aberrant comparison (OTU 27 and 31) of the 146 intraspecific comparisons in their study (Foley *et al.*, 1993: table 3). Its removal reduces the level of intraspecific variation to 0–14% and 0.007–0.267 Nei's D. Foley *et al.* (1994) found that fixed differences within OTUs of the *An. punctulatus* complex in the Solomon Islands were never more than 12% and 0.169 Nei's D (Foley *et al.*, 1993: table 2). Thus, the levels of genetic divergence used in the present study for determining whether the hypothesis of conspecificity is rejected are conservative; separate species can differ by less than these levels and remain undetected but it is unlikely that groups that differ by more than these levels are conspecific.

Genetic distances were clustered using the UPGMA and NJ algorithms in MEGA, version 3 (Kumar, Tamura & Nei, 2004). The UPGMA assumes that the rate of evolution has remained constant throughout the evolutionary history of the included taxa, and thus a rooted tree is produced. The NJ method (Saitou & Nei, 1987) produces an unrooted tree because it does not require the assumption of a constant rate of evolution. MEGA provides an option of a linearized version of the NJ tree, which assumes a constant rate of evolution. The number of species was estimated by inspection of the UPGMA and linearized NJ trees.

Alternative UPGMA trees can be produced from the same data by different computer programs and because of data input order effects (i.e. ties) (Backeljau *et al.*, 1996). By comparison, the NJ algorithm does not force sister OTUs to display equal branch lengths and tie trees are rare for this method (Takezaki, 1998). A measure of the robustness of trees was obtained by the program TFPGA 1.3 (Miller, 1997), which can produce Bootstrap values (100 replicates) for UPGMA trees based on Nei's D (Nei, 1978) corrected for small sample size. For convenience, bootstrap values were displayed on a tree produced in MEGA rather than TFPGA because the latter can draw UPGMA dendrograms incorrectly if there are tied trees (Miller, 1997).

MODEL-BASED ANALYSIS

We ran STRUCTURE with the non-admixture model of ancestry plus the option of uncorrelated allele frequencies, and the admixture model plus the correlated allele frequency option. The former settings are appropriate for very discrete populations (Pritchard & Wen, 2003). Although the non-admixture settings appear *a priori* to be most appropriate for species level comparisons, D. H, Foley (unpubl. data) found that the admixture settings resulted in better estimates of the correct species number. The non-admixture model assumes the allele frequency of each population is an independent draw from a distribution specified by λ, which for the Australia-wide analysis was estimated to be 0.4183 for $K = 1$ and was fixed at this level thereafter, as recommended in the STRUCTURE manual. For the admixture model, λ was set to one, as the manual advises. Burn-in was set at 10 000 and Markov Chain Monte Carlo at 50 000 for at least ten replicates up to $K = 40$. In addition to an Australia-wide analysis, regional-based analyses were performed for QLD, NSW plus TAS plus VIC, and WA plus NT.

As the STRUCTURE algorithm sometimes converges towards modes of much lower likelihood (i.e. multimodality of Pritchard & Wen, 2003), we followed the method for identifying and replacing outliers as proposed by D. H. Foley (unpubl. data). Briefly, we characterized the degree of asymmetry of the distribution of replicate $\ln(K)$ values around the mean for a given K, using the Skewness function in Microsoft Office Excel 2003 (Microsoft Corp.). Skewness values less than -1 were identified and the lowest $\ln(K)$ values removed until skewness was greater than -1. The scatter of the points was inspected and if sharply defined multiple modes were present, the lower probability points were removed. We calculated the *ad hoc* quantity (ΔK) (Evanno, Regnaut & Coudet, 2005) to assist the identification of the actual number of groups. Evanno *et al.* (2005) used the height of the modal value of ΔK as an indicator of the strength of the signal detected by STRUCTURE.

Under admixture settings, an individual was assigned to a cluster if its membership value for that cluster was ≥ 0.500, if the value was less than 0.500, the individual's assignment was treated as unknown.

DISTRIBUTION MODELLING

The potential distribution of clusters identified by the STRUCTURE analysis was predicted using BIOCLIM (Nix, 1986) in DIVA-GIS 4.1. BIOCLIM attempts to identify suitable and unsuitable areas or 'niches' in which the organism could occur based on the climatic and ecological features of the sampled data points. The BIOCLIM model was implemented using the True–False option and the WORLDCLIM 2.5-min resolution database of 19 bioclimatic variables (i.e. annual mean temperature, mean monthly temperature range, isothermality, temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, annual temperature range, mean temperature of the wettest quarter, mean temperature of the driest quarter, mean temperature of the coldest quarter, mean temperature of the warmest quarter, annual precipitation, wettest month precipitation, driest month precipitation, precipitation seasonality, wettest quarter mean precipitation, driest quarter mean precipitation, coldest quarter mean precipitation, and warmest quarter precipitation).

RESULTS

A total of 366 specimens of *An. annulipes s.l.* from 61 sites (Table 1) were subjected to electrophoresis. Samples were scored for up to 32 putative allozyme loci (i.e. *Acon-1*, *Acon-2*, *Acp*, *Ak-2*, *αAmy*, *Enol*, *Fdp-2*, *βGal-1*, *βGal-2*, *Got-1*, *Got-2*, *αGpd*, *6-Gpd*, *Gpi*, *Hbdh*, *Hk-1*, *Hk-2*, *Hk-3*, *Idh-1*, *Idh-2*, *Ldh*, *Mdh-1*, *Me-1*, *Mpi-2*, *Odh*, *PepB-1*, *PepB-2*, *PepD-1*, *PepD-2*, *Pgm*, *Pk*, and *Thdh*) containing up to nine alleles. Individuals of *An. annulipes s.l.* sorted into 152 OTUs (Table 1) and the genetic profile of groups of OTUs that differed by less than 15%FD are shown in the Appendix.

Outgroups comprising other species within the subgenus *Cellia* from Australia (i.e. *An. farauti*, *An. torresiensis*, *An. hinesorum*, *Anopheles amictus* Edwards, *Anopheles hilli* Woodhill & Lee, *Anopheles meraukensis* Venhuis, *Anopheles novaguinensis* Venhuis) formed a cluster separate to *An. annulipes s.l.* (data not shown). This suggests that *An. annulipes s.l.* is monophyletic and confirms that specimens were correctly assigned to this taxon. Figure 1 shows the NJ tree of %FD. The UPGMA of Nei's *D* (Fig. 2) resulted in three tied trees and all of the 100 bootstrap trees resulted in tied trees. Misleading results are likely when tied trees exist and when a high proportion of bootstrap replicates result in the formation of tied trees (Backeljau *et al.*, 1996). The presence of tied trees and the observation that most branches had low bootstrap support (Fig. 2) indicates that alternate topologies exist to the topology of the tree shown here.

STRUCTURE ANALYSIS

Only 28 loci were included in the STRUCTURE analysis of Australia-wide data; *Got-1*, *Hk-1*, *Hk-2*, and *Hk-3* were excluded because of a lack of variability. The output for the admixture and non-admixture settings is given in Figure 3. The strongest peaks in ΔK for the admixture model were for $K = 2$, followed by $K = 4$, $K = 7$ and then $K = 20$. Mean $\ln(K)$ starts to plateau between $K = 15$–20, which suggests that any structure identified by ΔK below this level is spurious or due to supraspecific structure. Maximum $\ln(K)$ was between $K = 20$–25, with the lowest maximum $\ln(K)$ (i.e. highest probability) score for the entire admixture modelling at $K = 25$. High variation in $\ln(K)$ beyond $K = 25$, even with outlying low likelihood scores removed, makes it difficult to discern the extent of the plateau. The strongest peak in ΔK for the non-admixture model was for $K = 2$ followed by $K = 7$, and then $K = 12$. Mean and maximum $\ln(K)$ did not plateau but continued to increase up to $K = 40$, the highest K analysed. D. H. Foley (unpubl. data) found that the admixture settings resulted in a stronger species signal (e.g. height of ΔK), and the lack of a plateau for non-admixture settings in the present study confirms that the admixture settings are more appropriate for complex allozyme data sets.

The distribution of clusters among the sample sites for $K = 15$, 20 and 25 under the admixture settings is shown in Table 1 and Figure 4. In many locations, the numbers of clusters did not change with increasing K; for example, the number of clusters at Clark River remained at 4 regardless of increasing K from 15 to 25 (Table 1). In other cases, the number of clusters increased or (more rarely) decreased with increasing K. The results from the STRUCTURE analysis for $K = 15$–25 indicated that most (> 94%) OTUs were assigned to one or other cluster but that the numbers that were split among different cluster were slightly higher for non-admixture than admixture settings. The number of specimens that could not be assigned to clusters increased with K (for $K = 15$, 20 and 25) and was greater for admixture ($N = 8$, 17, 17) than non-admixture ($N = 5$, 11, 7) settings.

NUMBER OF SPECIES

Table 2 shows the number of species estimated from the tree-based analyses using the 20%FD and 0.300 Nei's *D* levels for rejection of conspecificity. The number of species is in the range 18–25 for the Australia-wide analyses. The Nei's *D* intraspecific threshold resulted in a more conservative estimate of species than did the %FD threshold. The NJ algorithm was more conservative than UPGMA, especially for %FD.

The numbers of clusters that occur in geographical regions within Australia are also shown in Table 2

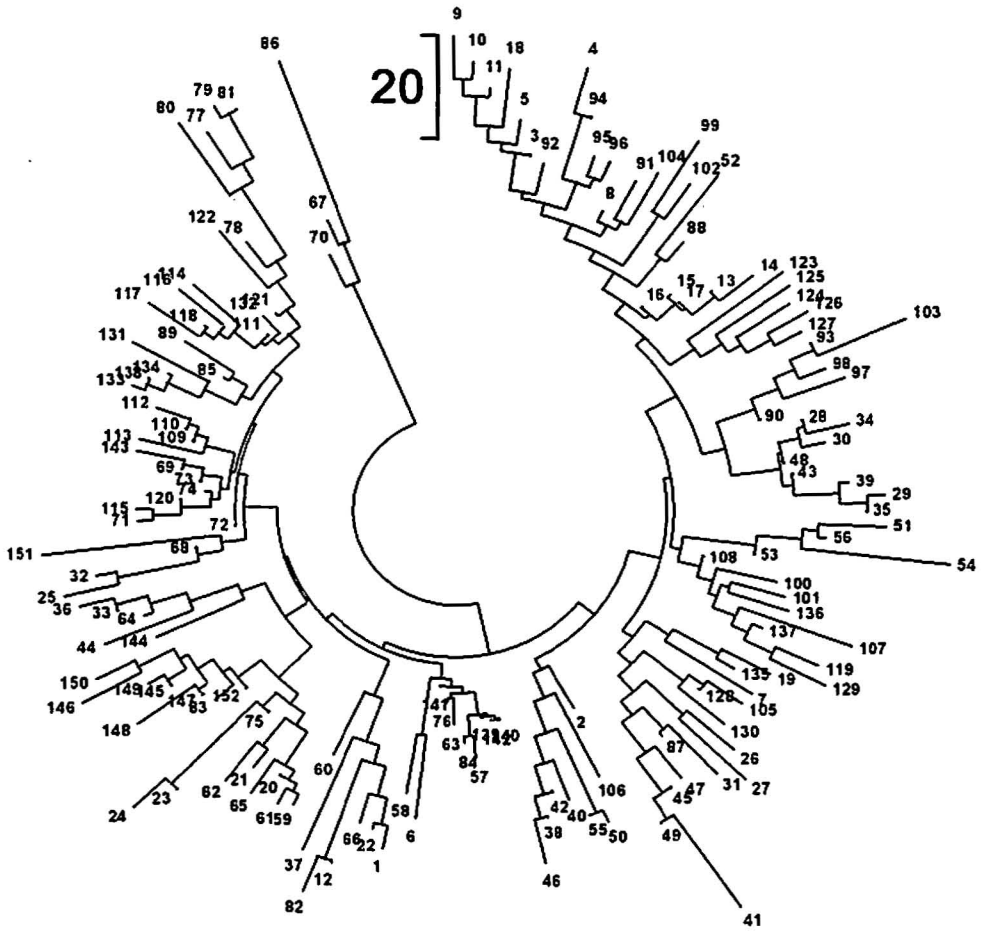

Figure 1. Neighbour-joining tree of percent fixed differences (%FD) for allozymes of 152 operational taxonomic units (OTUs) of *Anopheles annulipes s.l.* For OTU information, see Table 1.

Table 2. Number of species within *Anopheles annulipes s.l.* estimated to occur according to Bayesian and tree-based clustering analysis of allozyme data

	Bayesian				UPGMA*		NJ†	
Geographic distribution	Total	*K*15	*K*20	*K*25	%FD	Nei's *D*	%FD	Nei's *D*
WA + NT	7–11	(6)	(11)	(13)	9 (10)	7 (9)	8 (10)	8 (9)
NSW + TAS + VIC	9	(7)	(9)	(11)	9 (10)	8 (9)	7 (10)	6 (9)
QLD	10	(11)	(12)	(18)	14 (15)	11 (12)	12 (14)	11 (12)
Australia	15–25	15	20	25	25	18	24	18

The estimated number of species is shown for the analysis of combined data (Australia) and for analysis of data from different geographical regions. The geographical distribution of species number according to the results of the Australia-wide analysis is shown in brackets. The species number by geographical region from the Bayesian analysis is also shown assuming 15, 20 and 25 groups (*K*) within Australia.

Numbers of species determined at 20%FD and 0.300 Nei's *D* (corrected for small sample size) according to *unweighted pair group method of analysis (UPGMA) and †Neighbour-joining (NJ) algorithms.

WA, Western Australia; NT, Northern Territory; NSW, New South Wales; TAS, Tasmania; VIC, Victoria; QLD, Queensland.

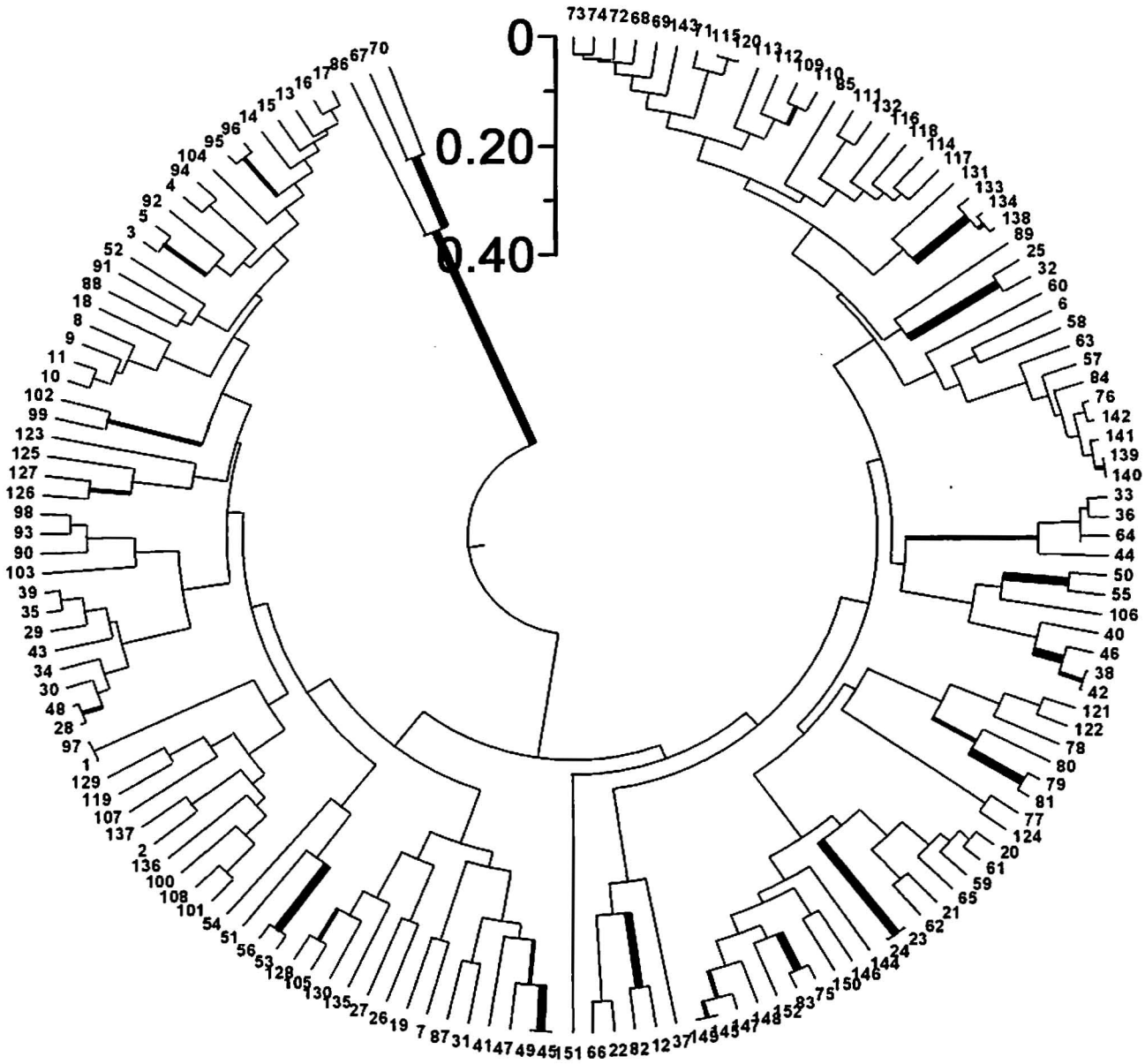

Figure 2. Unweighted pair group method of analysis (UPGMA) tree of Nei's *D* for allozymes of 152 operational taxonomic units (OTUs) of *Anopheles annulipes s.l.* Branches with thick lines indicate > 50% and thickest lines > 70% bootstrap support (100 replicates). For OTU information, see Table 1.

according to Bayesian clustering at $K = 15$, 20 and 25 for the Australia-wide analysis, and for separate tree-based analyses. Tree-based estimates of species number occurring within these geographical subregions generally coincide with model based estimates for $K = 15$–20. However, the Australia-wide comparison suggests that the number of species is higher at $K = 20$–25. The reason for this discrepancy is not known but may be attributable in part to the larger and more complex Australia-wide data set compared with those for subregions. Thus, a conservative estimate of the number of species from the STRUCTURE analysis is 15–20. The lower number is equivalent to 25%FD and 0.310 Nei's *D*; higher than the intraspecific threshold indicated by the *An. punctulatus* group data.

DISTRIBUTION

The geographical distribution of the $K = 15$ clusters is shown in Figure 4 and smaller maps show the predicted distribution of individual clusters based on out-

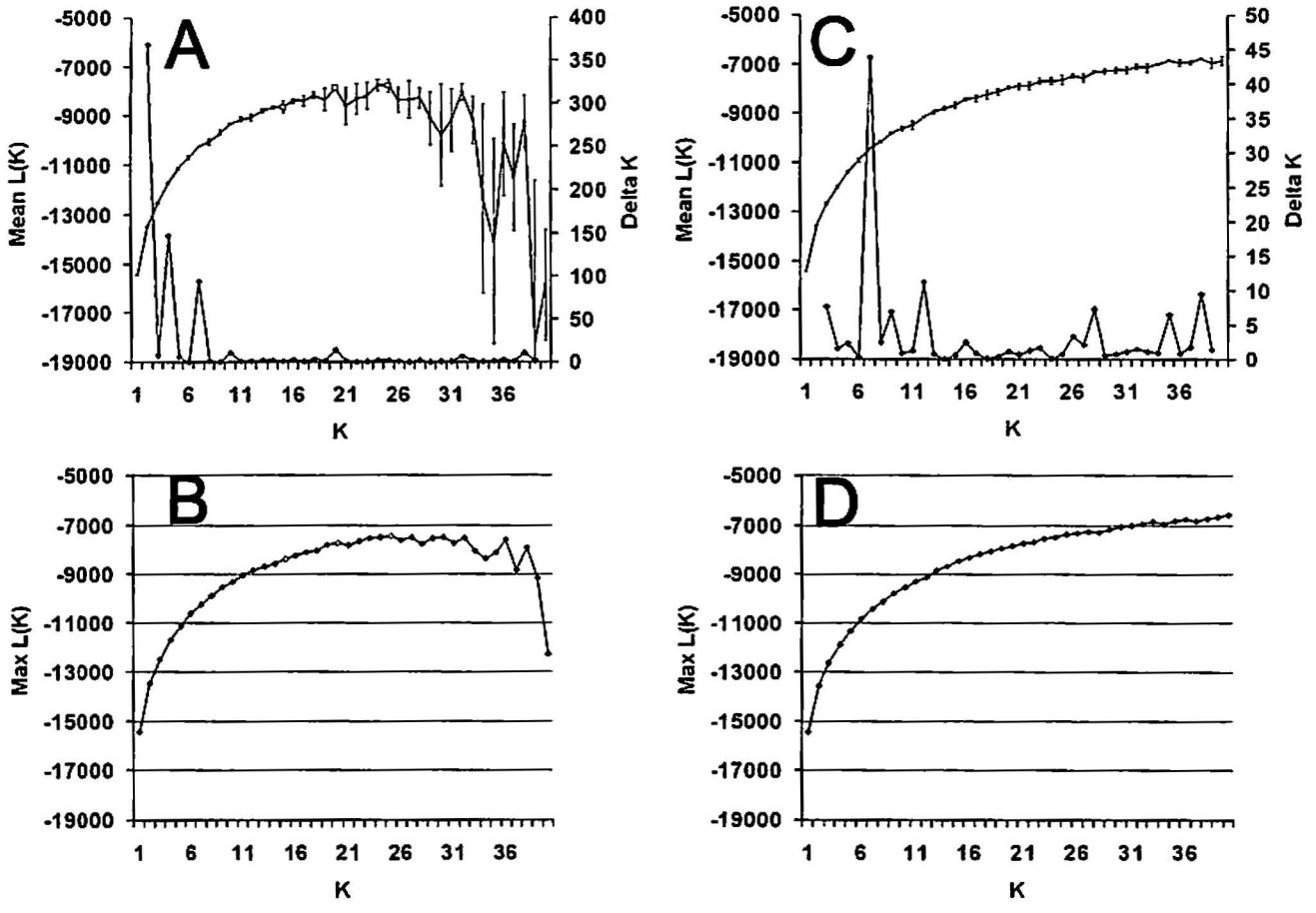

Figure 3. A, B, C, D, mean ± standard deviation ln(K), Maximum ln(K) and ΔK for 10–22 STRUCTURE runs for $K = 1$–40 for *Anopheles annulipes s.l.* allozyme data under admixture settings (A, B) and 10–33 runs for non-admixture settings (C, D). The modal value of the distribution of ΔK is intended to indicate the true K, or the uppermost level of structure. $K = 15$, 20 and 25 under admixture settings are given with outline symbol. $K = 2$ for non-admixture run ($\Delta K = 925$) is not shown.

put from the BIOCLIM true–false ecological niche model. The number of input locations for some clusters was very low and the predicted distribution is not shown for these. For presentation purposes, clusters whose predicted distribution was of limited geographical extent also are not shown. For example, cluster 7 was predicted to occur only in isolated points along coastal northern NSW and southern QLD, and cluster 13 along coastal south and central NSW and eastern VIC. Although New Guinea was included in the area encompassed by the modelling, no clusters were predicted to occur there.

ALLOZYME VARIATION

The distribution of genotypes of *Gpi* according to latitude (decimal degrees south) is shown in Figure 5. Alleles 1–4 are shown on the y-axis, in order of mobility, with the slowest allele numbered '1'. Hybrids of consecutive alleles also are shown (e.g. genotype 1,2 is shown as 1.5). Only three genotypes were hybrids of nonconsecutive mobility alleles and are not shown. A clear trend from slow mobility alleles in the north to fast alleles in the south can be seen. Individuals are shown in Figure 5 according to their membership of $K = 2$ clusters from the STRUCTURE analysis under admixture settings. Specimens of cluster 1 are more likely to occur in the north and have slow alleles whereas specimens of cluster 2 are more likely in the south and have fast alleles.

DISCUSSION

The present allozyme study attempts to determine how many species may occur within the *An. annulipes* complex by comparing tree-based approaches for clustering OTUs with a model-based Bayesian approach for clustering individual genotype data. Recently, a

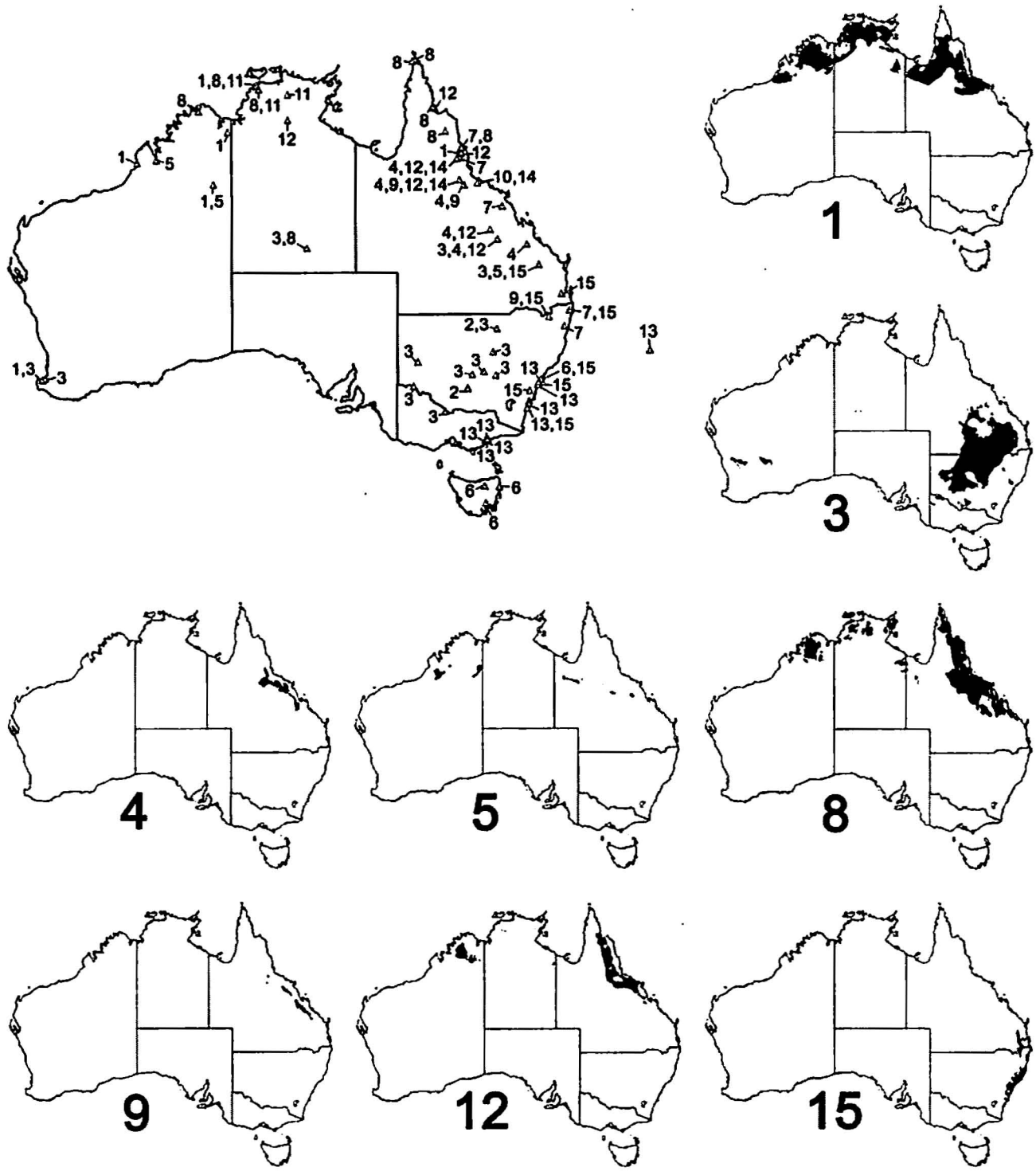

Figure 4. Map of Australia with the location of $K = 15$ clusters identified from a STRUCTURE analysis of allozyme genotype data for 366 *Anopheles annulipes s.l.* Smaller maps show the predicted distribution of a selection of individual clusters based on output from the BIOCLIM true–false ecological niche model available in DIVA-GIS. *Anopheles annulipes* sp. A = cluster 3, sp. B. = 8, sp. C = 13, sp. D = 1, sp. E = 15, sp. F = 7, sp. G = 2, and Mataranka chromotype = 12.

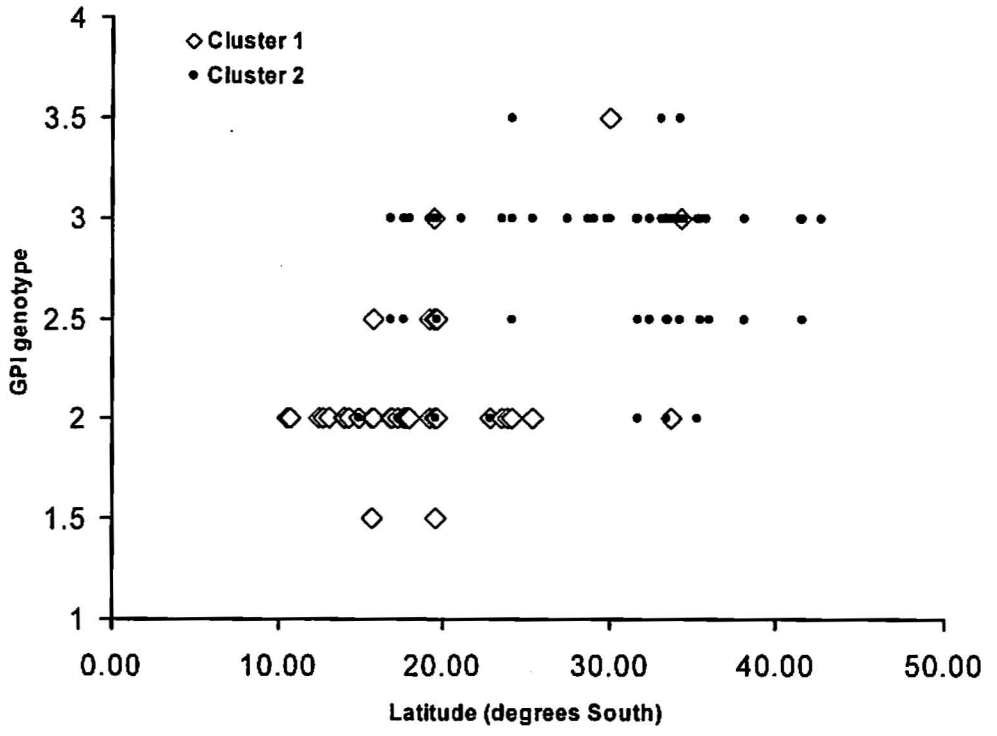

Figure 5. The distribution of genotypes of *Gpi* of *Anopheles annulipes s.l.* according to latitude (decimal degrees south) and membership of $K = 2$ clusters from a STRUCTURE analysis carried out under admixture settings. Alleles 1–4 are shown on the *y*-axis, in order of mobility, with the slowest allele numbered '1'. Hybrids of consecutive alleles are also shown (e.g. genotype 1,2 is shown as 1.5).

reanalysis of allozymes of the *An. punctulatus* group using the Bayesian clustering approach implemented in the program STRUCTURE successfully identified the correct number of species, suggesting a new approach for determining the number of species in a sample of genotypes (D. H. Foley, unpubl. data).

Both the Bayesian and tree-based clustering approaches indicate that the *An. annulipes* complex contains more species than previously suspected. However, the Bayesian approach may be more reliable because the STRUCTURE algorithm was explicitly designed to overcome the limitations of genetic distance matrix-based methods, which lose information through collapsing genotype data for pairs of species into single numbers (Pritchard, Stephens & Donnelly, 2000). The Bayesian approach also may give a more accurate estimate of species number than tree-based approaches if the evolutionary history of *An. annulipes s.l.* is not well represented by a bifurcating tree.

For tree-based methods, the Nei's *D* species cut-off values gave more conservative estimates of species number than %FD. The frequency of tied trees and the amount of OTU clustering inconsistency between UPGMA and NJ trees should be higher if the algorithm (especially UPGMA) is forced to display OTUs that differ by the same amount. The average %FD for OTUs in the Australia-wide analysis was 22.23 ± 10.85, compared to 49.3 ± 22.06 for the *An. punctulatus* group calculated from the data set of Foley *et al.* (1995). This difference probably reflects the greater evolutionary divergence and accumulation of interspecific allozyme differences within the *An. punctulatus* group compared with species within the *An. annulipes* complex. The narrower range of genetic distances in the Australia-wide analysis of the *An. annulipes* complex may have inflated the estimates of species number compared to those from separate analyses of geographical subregions.

The present study uses a phenetic rather than a phylogenetic approach to species delineation. We assume that historical and contemporary gene flow between individuals of a species will limit genetic divergence within species compared with divergence between most species. This species signal can be seen most clearly in sympatric locations by observation of a lack of hybridization indicating the presence of two or more species. For comparison of allopatric mosquito populations, assortative mating cannot be observed but genetic divergence can be measured and individual genotypes clustered accordingly. All of the previously recognized sibling species of *An. annulipes s.l.*

included in the present study cluster as separate species according to the tree and model-based approaches. Thus, we assume that the species-level clusters that we identify represent real biological species that will not hybridize in sympatry and have independent evolutionary histories, and possibly marked differences in biology and behaviour.

The present study suggests that at least 15–20 species are represented among the specimens analysed. Previous estimates, based on crossmatings and polytene chromosome analysis of a smaller number of specimens, suggested at least ten sibling species (Booth & Bryan, 1986). *Anopheles annulipes s.l.* appears to be the most species-rich anopheline species complex known to date; the *Anopheles gambiae* complex has eight species, and the *Anopheles crucians*, *An. farauti* and *Anopheles dirus* complexes each have seven species (Harbach, 2004).

The reason that *An. annulipes s.l.* has undergone such an extensive species radiation is unknown. Latitude-dependent variation was found for the *Gpi* locus of *An. annulipes s.l.* Populations of *Colias* butterflies with different alleles of *Gpi* vary in dispersal ability and fitness according to ambient temperature and elevation (Watt *et al.*, 2003). Temperature in particular can influence the functioning of enzymes, which, in the case of *Pgm* and the yellow dung fly *Scathophaga stercoraria* (L.), may determine the outcome of sexual selection (Ward, Jann & Blankenhorn, 2004). The apparent cline in *Gpi* for the *An. annulipes* complex suggests a similar influence of temperature on distribution. STRUCTURE analysis can detect supraspecific phylogenetic groupings (D. H. Foley, unpubl. data) and the high ΔK for $K = 2$ suggested two clades within the *An. annulipes* complex. The STRUCTURE analysis and the geographical distribution of OTUs according to tree-based clustering suggest that *An. annulipes s.l.* is composed of a similar number of northern and southern species. Thermal, or some other latitude-dependent adaptation, may have had an important role in speciation and the subsequent distribution of the *An. annulipes* complex. Foley, Russell & Bryan (2004) noted that north Australian *Ochlerotatus notoscriptus* (Skuse) also possessed unique slow mobility alleles of *Gpi*.

The identity of some specimens was suspected based on reports of the geographical distribution of chromosomally identified forms. In 1977, C. A. Green (unpubl. data) identified sp. A, sp. B, sp. C, and sp. D from sites throughout Australia based on chromosomes. Booth, Green & Bryan (1987) showed a distribution map of chromosomally identified species for Australia (Booth *et al.*, 1987: fig. 3). *Anopheles annulipes* sp. A or sp. G were suspected from Griffith and Hanwood, NSW based on the allozyme study of Foley & Bryan (1991a). From the chromosomal identity and distribution of $K = 15$ clusters, the tentative identification of species is: *An. annulipes* sp. A = cluster 3, sp. B = 8, sp. C = 13, sp. D = 1, sp. E = 15, sp. F = 7, sp. G = 2, and Mataranka chromotype = 12. The distribution of the 20 clusters identified by the STRUCTURE analysis and the matching of clusters with chromosomal types reported in the present study is largely concordant with the reported distribution of these types, although important differences occur. C. A. Green (unpubl. data) and Booth & Bryan (1986) reported sp. A from the type locality (TAS) but we found only one cluster (18) in 15 specimens from four sites in TAS, which did not match sp. A (clusters 10 and or 14). Further sampling within TAS may reveal the presence of more sibling species.

Liehne (1991) states that *An. annulipes* sp. D is found in northern WA. Specimens from Alice Springs, NT that conformed to his description of this species were included in the molecular phylogeny of Foley *et al.* (1998). However, the allozyme cluster that is most common in northern WA was not found in Alice Springs. It is possible that *An. annulipes* sp. D occurs in Alice Springs but was not sampled in the present study or that the specimen used by Foley *et al.* (1998) was another, as yet undetermined, species.

From the distribution of clusters, the syntypes from Sydney could have included sp. C, sp. E, and cluster 18. Although the localities of the other syntypes (i.e. Adelaide River, NT and Irvinebank, QLD) were not sampled, specimens from nearby sites indicate that a number of clusters are candidates. Inferences about the identity of types will be problematic, especially for the Sydney specimens, due to the amount of environmental modification at the site and lack of details about the Sydney and Tasmanian locations.

The predicted distribution of each of $K = 15$ clusters did not extend to New Guinea despite the presence of *An. annulipes s.l.* there. It is likely that *An. annulipes s.l.* in New Guinea consists of sibling species that were not sampled in this study. The presence of sp. C on Lord Howe Island is likely to be the result of an introduction from populations from coastal NSW.

The ecological niche modelling conducted in the present study was an attempt to gain insight into gross differences in the potential distribution of sibling species of *An. annulipes s.l.*, and not to comprise a definitive prediction of distribution. A better assessment of potential distribution will require greater sampling, preferably of molecular-typed specimens, and a statistical treatment of the reliability of distribution models, as is available in the modelling procedure Genetic Algorithm for Rule Set Prediction (Stockwell & Noble, 1992) available in the DESKTOP GARP software.

Anopheles annulipes s.l. is the most important vector of myxomatosis in many areas of Australia (Fenner

& Ratcliffe, 1965; Parer & Korn, 1989), and the possibility of more than one biological form of *An. annulipes* has been suggested to explain geographical differences in the ability of the myxoma virus to control rabbits (Fenner & Ratcliffe, 1965). Although an epidemiological assessment of the role in myxomatosis transmission of the different sibling species will have to wait a more detailed survey, some preliminary observations can be made. Lee, Clinton & O'Gower (1954) noted that *An. annulipes s.l.* from river flats at Merbein, VIC predominantly fed on rabbits despite the rabbit population having been decimated by myxomatosis. As Merbein is near Mildura, VIC where *An. annulipes* sp. A was identified, it is likely that this species was among those surveyed by Lee *et al.* (1954). Fenner & Ratcliffe (1965) matched the presence of *An. annulipes s.l.* with an epizootic of myzomatosis at Yarram, VIC and, according to our study *An. annulipes* sp. E occurred close to this site.

Additional cryptic species may await detection. For example, although a form of *annulipes* with a black proboscis is known (e.g. *An. musivus*), such specimens may have been omitted from our sample as they could be confused with other species of the subgenus *Cellia*. Missing data and genotyping errors are likely to have contributed to the complexity of the data set, which could have affected estimates of species number. For example, although critical side-by-side comparisons of bands (i.e. line-ups; Richardson *et al.*, 1986) are possible with the electrophoresis system used in the present study (Foley, 1990), the number of line-ups was limited by the amount of sample that could be obtained from one mosquito. For the input to STRUCTURE, 10.8% of data were missing. However, despite a higher level of missing values, D. H. Foley (unpubl. data) was able to reveal the correct species composition of the *An. punctulatus* group, suggesting that the Bayesian approach is robust and reliable. The low branch support and the presence of tied trees found in the present study indicate that alternative topologies exist, but the approximate agreement in species number estimated by model and different tree-based approaches suggests that this estimate is relatively unaffected by branch instability.

Recently, molecular markers have replaced allozymes for population and species studies. However, the Bayesian approach used in the present study offers a new and powerful way for analysing multilocus genotype data, including DNA-based and allozyme data.

ACKNOWLEDGEMENTS

This project was not possible without specimens kindly supplied by the following: Tony Barnes, Ann Marie Boyd, Bob Cooper, Peter Ebsworth, Ian Fanning, Chris Freebairn, Brian Kay, Con Lokkers, Melina Miles, Andrew Van Den Hurk, Gimme Walter (QLD); David Booth, Stephen Doggett, Merelyn Geary, Richard Russell, Margaret Spencer (NSW); Brian Cordwell, Gary George, Michael Holland, M. Kennedy, Brian Watson (VIC); Tony Wright (WA); Tom Burkot, Jim Stott, Peter Whelan (NT); and Jo Kent and Craig Williams (South Australia). We thank Mark Adams of the South Australian Museum for advice and the use of unpublished computer programs for allozyme analysis. This research was completed while D.H.F. held a National Research Council Research Associateship Award at the Walter Reed Army Institute of Research. This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. The published material reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense.

REFERENCES

Backeljau T, De Bruyn L, De Wolfe H, Jordaens K, Van Dongen S, Winnepenninckx B. 1996. Multiple UPGMA and Neighbor-joining trees and the performance of some computer packages. *Molecular Biology and Evolution* **13:** 309–313.

Black RH. 1972. *Malaria in Australia.* Canberra: Australian Government Publishing Service.

Booth DR, Bryan JH. 1986. Cytogenetic and crossbreeding evidence for additional cryptic species in the *Anopheles annulipes* Walker complex (Diptera: Culicidae). *Journal of the Australian Entomological Society* **25:** 315–325.

Booth DR, Green CA, Bryan JH. 1987. The larval salivary gland polytene chromosomes of *Anopheles (Cellia) annulipes* s.l. Walker (Diptera: Culicidae). *Australian Journal of Zoology* **35:** 247–252.

Bryan JH, Foley DH, Geary M, Carven CTJ. 1991. *Anopheles annulipes* Walker (Diptera: Culicidae) at Griffith, New South Wales. 3. Dispersal of two sibling species. *Journal of the Australian Entomological Society* **30:** 119–121.

Evanno G, Regnaut S, Coudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Fenner F, Ratcliffe GN. 1965. *Myxomatosis.* Cambridge: Cambridge University Press.

Foley DH. 1990. An improved applicator system for cellulose acetate electrophoresis. *Journal of the American Mosquito Control Association* **6:** 141–143.

Foley DH, Barnes A, Bryan JH. 1992. *Anopheles annulipes* Walker (Diptera: Culicidae) at Griffith, New South Wales. 4. Phenology of Two Sibling species. *Journal of the Australian Entomological Society* **31:** 91–96.

Foley DH, Bryan JH. 1991a. *Anopheles annulipes* Walker (Diptera: Culicidae) at Griffith. New South Wales. 1. Two sib-

ling species in sympatry. *Journal of the Australian Entomological Society* **30:** 109–112.

Foley DH, Bryan JH. 1991b. *Anopheles annulipes* Walker (Diptera: Culicidae) at Griffith, New South Wales. 2. Biology and behaviour of two sibling species. *Journal of the Australian Entomological Society* **30:** 113–118.

Foley DH, Bryan JH, Yeates DK, Saul A. 1998. Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. *Molecular Phylogenetics and Evolution* **9:** 262–275.

Foley DH, Cooper RD, Bryan JH. 1995. A new species within the *Anopheles punctulatus* complex in Western Province, Papua New Guinea. *Journal of the American Mosquito Control Association* **11:** 122–127.

Foley DH, Meek SR, Bryan JH. 1994. The *Anopheles punctulatus* group of mosquitoes in the Solomon Islands and Vanuatu surveyed by allozyme electrophoresis. *Medical and Veterinary Entomology* **8:** 340–350.

Foley DH, Paru R, Dagoro H, Bryan JH. 1993. Allozyme analysis reveals six species within the *Anopheles punctulatus* complex of mosquitoes in Papua New Guinea. *Medical and Veterinary Entomology* **7:** 37–48.

Foley DH, Russell RC, Bryan JH. 2004. Population structure of the peridomestic mosquito *Ochlerotatus notoscriptus* in Australia. *Medical and Veterinary Entomology* **18:** 180–190.

Harbach RE. 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bulletin of Entomological Research* **94:** 537–553.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5:** 150–163.

Lee DJ, Clinton KJ, O'Gower AK. 1954. The blood sources of some Australian mosquitoes. *Australian Journal of Biological Sciences* **7:** 282–301.

Lee DJ, Hicks MM, Griffiths M, Debenham ML, Bryan JH, Russell RC, Geary M, Marks EN. 1987. *The Culicidae of the Australasian region*, Vol. 5. Canberra: Australian Government Publishing Service.

Liehne PFS. 1991. *An atlas of the mosquitoes of Western Australia*. Perth: Health Department of Western Australia.

Miller MP. 1997. *Tools for population genetic analysis*, Version 1.3. Flagstaff, AZ: Department of Biological Sciences, Northern Arizona University.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583–590.

Nix HA. 1986. A biogeographic analysis of Australian elapid snakes. In: Longmore R, ed. *Atlas of elapid snakes of Australia*. Australian Flora and Fauna Series no. 7. Canberra: Australian Government Publishing Service, 415.

Parer I, Korn TJ. 1989. Seasonal incidence of myxomatosis in New South Wales. *Australian Wildlife Research* **16:** 563–568.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Pritchard JK, Wen W. 2003. *Documentation for STRUCTURE software*, Version 2. Available at http://pritch.bsd.uchicago.edu.

Richardson BJ, Baverstock PR, Adams M. 1986. *Allozyme electrophoresis. A handbook for animal systematics and population studies*. Sydney: Academic Press.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Maki-Tanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159:** 699–713.

Russell RC. 1995. Arboviruses and their vectors in Australia: an update on the ecology and epidemiology of some mosquito-borne arboviruses. *Review of Medical and Veterinary Entomology* **83:** 141–158.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406–425.

Stockwell DRB, Noble IR. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* **32:** 249–254.

Takezaki N. 1998. Tie trees generated by distance methods of phylogenetic reconstruction. *Molecular Biology and Evolution* **15:** 727–737.

Ward PI, Jann P, Blanckenhorn WU. 2004. Temperature-mediated seasonal variation in phosphoglucomutase allozyme frequency in the yellow dung fly, *Scathophaga stercoraria*. *Molecular Ecology* **13:** 3213–3218.

Watt WB, Wheat CW, Meyer EH, Martin JF. 2003. Adaptation at specific loci. VII. Natural selection, dispersal and the diversity of molecular-functional variation patterns among butterfly species complexes (*Colias*: Lepidoptera, Pieridae). *Molecular Ecology* **12:** 1265–1275.

APPENDIX

Table A1. Allelic composition and numbers of specimens in 47 clusters of *Anopheles annulipes s.l.* comprising operational taxonomic units that exhibit less than 15% fixed differences from one another (*Got-1*, *Hk-1*, *Hk-2* and *Hk-3* not shown due to lack of variability)

	Locus																											
Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	b	c	d	ab	abc	bc	c	b	b	c	c	cd	ab	bc	c	c	fgh	b	c	cd	ef	c	bc	eg	cd	bc	ab	ab
	8	8	8	5	7	8	1	8	8	8	8	3	8	8	8	7	8	8	8	7	8	4	8	8	8	8	8	8
2	b	c	d	a	c	c	–	b	b	c	c	d	a	b	–	e	f	b	e	–	–	c	b	e	c	c	–	–
	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0	0
3	d	c	d	b	bc	b	b	b	b	c	c	cd	b	d	c	c	ab	b	e	g	f	bc	bc	df	c	c	b	–
	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	1	1	4	4	4	4	4	4	0
4	d	c	cd	b	abc	abc	c	b	b	c	c	c	b	d	d	bc	abf	b	c	a	ef	bce	c	egh	c	c	ab	ab
	6	6	6	6	6	6	4	5	6	6	6	4	6	6	6	6	6	6	6	2	6	6	6	6	6	6	6	6
5	d	c	cd	a	b	a	cd	b	b	c	c	c	b	b	c	c	a	a	eg	g	e	bce	c	dh	c	c	ad	–
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	0
6	d	c	bd	a	bc	bc	bcd	b	b	c	bc	c	a	b	b	c	cd	a	ce	–	de	b	bc	c	c	c	d	–
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	2	4	4	4	4	4	4	0
7	bd	c	d	a	c	a	–	b	b	c	bc	–	a	cd	d	b	bf	ab	c	deg	ef	bc	c	eg	cd	ac	ab	a
	2	2	2	2	2	2	0	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	d	c	d	b	bc	a	a	b	b	c	bc	cd	ab	d	b	c	cd	a	ce	bd	b	bc	ce	ce	c	c	abd	–
	7	7	7	7	7	7	6	7	7	7	7	7	7	7	6	7	7	7	7	2	2	7	7	7	7	7	7	0
9	d	c	cd	–	bc	a	a	b	b	c	c	cd	a	d	–	–	fg	–	–	d	b	c	bc	f	c	b	ad	–
	10	10	10	0	10	10	10	10	10	10	10	7	9	10	0	0	10	0	0	10	10	10	10	10	7	10	10	0
10	d	c	d	b	bc	a	bc	bd	b	bc	bc	cd	b	d	b	c	cd	ab	cde	g	de	ce	bc	b–fh	c	cd	a	–
	10	10	10	10	10	10	10	10	10	10	10	9	10	10	10	10	10	9	10	4	7	10	10	10	10	9	3	0
11	d	c	d	b	c	ab	ab	–	b	c	b	d	b	d	c	c	e	ab	c	–	e	c	c	df	c	cd	d	b
	2	2	2	2	2	2	2	0	2	2	2	1	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2
12	bd	c	d	b	bc	a	b	b	b	c	b	c	a	bc	c	c	d	ab	be	g	e	ce	bc	df	c	cd	ad	b
	4	4	4	4	4	4	4	4	4	4	4	2	4	4	2	4	4	4	4	3	4	4	4	4	4	4	3	1
13	b	c	d	–	c	c	b	b	c	ce	c	c	a	e	c	–	g	a	–	gh	e	d	ce	h	c	d	ad	–
	2	2	2	0	2	2	2	2	2	2	2	2	1	2	1	0	2	2	0	2	2	2	2	2	2	2	2	0
14	bd	ac	bd	–	bc	c	b	b	b	bce	bc	c	ab	bd	c	–	fg	a	–	gh	de	ce	abce	fhi	c	b	d	–
	10	10	10	0	10	10	10	10	10	10	10	10	10	10	4	0	10	5	0	10	10	10	10	10	10	10	10	0
15	b	c	d	–	bc	b	b	b	b	e	b	c	a	d	c	–	g	–	–	h	e	bce	c	fh	c	a	d	–
	4	4	4	0	4	4	4	4	4	4	4	4	3	4	4	0	4	0	0	4	4	4	4	3	4	4	3	0

APPENDIX *Continued*

Cluster	Locus 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
16	b	c	b	–	c	b	b	b	b	c	b	c	a	d	–	–	f	–	–	h	e	c	bc	fh	c	a	a	–
	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1	1	1	0
17	b	c	cd	b	bc	bc	bc	b	b	c	bc	cd	a	d	b	c	bc	a	c–f	dg	ef	ce	bc	b–e	c	abc	ad	–
	16	16	16	16	16	16	16	16	16	16	16	14	15	16	16	16	16	14	15	4	8	16	16	15	16	16	14	0
18	b	ac	bcd	bc	ac	ab	bc	b	b	bce	b	cd	a	b–e	bcd	bc	bcf	a	ce	f	de	cde	bc	cdeg	cd	ac	ad	ab
	22	24	24	18	23	24	9	8	24	24	24	20	24	24	14	24	24	24	16	2	24	24	24	24	16	23	23	24
19	b	c	c	a	c	b	c	b	b	–	b	d	a	d	c	b	c	a	e	g	e	c	b	–	c	a	a	–
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
20	b	c	cd	ab	bc	abc	bc	b	ab	bce	ab	cd	ab	cd	c	c	bcf	a	bce	g	de	ce	abce	b–e	c	ab	abd	a
	16	18	18	15	17	18	16	14	18	16	18	9	18	18	15	16	18	17	14	16	18	8	14	18	17	18	18	14
21	bd	c	cd	b	ab	b	b	b	b	ce	bcd	c	ab	cd	d	c	b	b	c	ab	ef	–	e	egh	d	b	ab	b
	5	5	5	4	5	4	2	5	5	5	5	3	5	5	5	5	5	5	3	3	5	0	4	5	5	5	5	5
22	e	ac	bd	b	abc	abc	ab	c	c	c	cd	be	a	bc	c	cd	fg	ab	c	bd	def	ab	ce	b–e	c	bc	ad	b
	8	8	8	4	8	7	6	8	8	8	8	5	8	8	6	8	8	6	7	2	7	8	6	8	7	8	7	2
23	d	c	cd	a	c	b	c	–	b	c	c	c	ab	b	c	c	c	a	c	–	d	b	bc	–	c	c	d	b
	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1
24	b	c	d	–	b	c	–	b	b	c	c	–	a	b	c	c	h	b	c	d	e	–	b	bf	d	b	ab	b
	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
25	b	c	cd	a	bc	bc	c	–	b	c	c	c	b	b	c	c	cf	b	c	–	e	c	b	g	c	c	ab	b
	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	0	1	2	2	1	2	2	2	2
26	d	abc	bcd	a	abc	ab	c	ab	b	bc	b–e	cd	abd	abc	cd	bc	ab	ab	ce	bcd	bde	bc	bcef	cdeg	bcd	bcd	ad	abc
	27	27	27	15	23	27	4	27	27	24	27	24	26	25	26	26	27	27	25	27	26	23	27	24	27	27	26	27
27	c	–	b	a	b	a	c	b	b	c	c	–	b	b	d	–	f	b	e	b	d	c	c	e	–	c	d	b
	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1
28	d	c	cd	b	bc	ab	b	b	b	c	bc	c	ab	bc	d	bc	fgh	a	ac	cdgh	ef	abc	ce	cef	cd	bc	abd	b
	22	23	23	8	21	23	23	23	23	23	23	10	21	10	23	23	20	23	15	16	23	20	23	23	18	23	14	23
29	d	c	cd	ab	bc	a	bc	b	b	c	c	c	a	b	d	b	f	a	ac	d	e	ace	bce	de	c	bc	abd	b
	8	8	8	7	8	8	8	8	8	8	8	6	8	8	8	8	8	8	4	3	8	8	8	7	4	8	8	8
30	bd	c	d	ab	c	c	c	b	b	c	c	–	b	b	d	b	f	a	c	–	e	c	bc	e	b	c	a	b
	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
31	d	c	d	b	c	b	b	b	b	e	bc	c	b	d	d	–	–	a	e	a	e	bc	ce	ce	c	c	d	b
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
32	d	c	d	bc	c	b	bc	b	b	bc	c	–	b	d	de	–	f	ab	ce	b	de	bc	abc	e	–	c	d	b
	3	3	3	2	3	3	3	3	3	3	3	0	3	3	3	0	3	3	3	1	3	3	3	1	0	3	3	3
33	d	c	cd	abd	bc	abc	bc	b	b	bc	c	c	bc	d	d	bc	bd	b	c	abd	f	bc	cd	ghj	c	c	ab	ab
	12	12	12	11	12	12	11	9	12	12	12	9	12	12	12	11	12	12	12	8	12	12	12	12	12	12	12	12

34	bd	c	cd	bc	bc	bc	abc	b	b	abce	c	c	ab	d	d	bc	bcdf	ab	bce	ab	d–g	bc	abce	e–h	cd	bc	ab	ab
	24	24	24	22	24	22	20	24	23	24	23	17	24	24	22	9	24	24	18	16	24	23	24	24	14	24	20	24
35	d	c	cd	a	abc	a	bc	b	b	bce	bc	c	ab	bc	d	bc	ab	ab	bc	bde	bde	bce	ce	eg	cd	bc	abd	ab
	7	7	7	6	7	7	6	7	7	7	7	2	7	7	5	7	7	7	6	6	7	7	7	6	7	7	7	4
36	d	c	bcd	ab	bc	bc	bcd	b	b	c	bc	bcd	abd	bc	d	bce	abc	ab	ce	bde	a–e	c	acef	b–h	cd	bcd	abd	ab
	24	24	24	6	22	24	22	24	24	24	24	23	23	24	24	21	24	24	22	24	22	24	24	22	24	24	24	17
37	d	c	d	b	a	b	c	b	b	c	c	cd	a	b	d	c	b	a	c	d	–	c	ce	d	d	b	bd	–
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0
38	b	c	d	–	ab	a	c	b	b	f	c	cd	a	b	d	ac	b	a	c	d	de	c	c	–	ad	c	d	–
	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
39	b	c	cd	–	bc	a	c	b	b	c	bc	c	b	b	ad	–	c	ab	ce	d	de	c	–	cg	cd	bc	abd	ab
	10	10	10	0	10	9	9	9	10	10	10	8	10	10	10	0	10	10	7	10	10	10	0	6	10	10	10	10
40	d	c	d	–	c	b	–	b	b	c	c	d	a	d	c	e	f	a	c	d	be	c	c	de	c	c	ab	a
	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
41	abd	ac	d	ab	abc	bc	bc	b	b	ce	bc	cde	ab	cd	cd	bce	bcef	a	bc	deg	de	c	abce	ef	cd	c	a–d	ab
	15	17	17	8	16	17	13	17	16	17	17	7	17	15	16	11	17	17	14	14	16	17	12	17	17	17	17	13
42	bd	c	d	ab	bc	ab	bc	b	b	cd	ab	d	b	cd	cd	b–e	cdf	ab	b–e	g	ef	bce	bce	bf	c	cd	abd	ab
	14	14	14	5	14	14	10	11	14	14	14	3	14	14	12	8	14	14	12	12	14	14	9	14	14	14	14	14
43	a	c	d	–	c	b	c	b	b	c	b	d	a	b	c	c	b	a	c	–	–	c	ab	d	d	c	a	–
	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0
44	b	c	d	–	c	b	c	b	b	c	b	d	a	de	c	c	f	b	c	g	e	c	bc	d	d	c	b	–
	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
45	bd	c	d	ab	abc	b	c	b	b	c	b	d	ab	de	c	c	ce	ab	ce	dg	e	c	c	abd	d	d	ad	–
	6	6	6	2	6	6	4	6	6	6	6	2	6	6	6	5	6	5	6	5	5	6	6	6	6	6	6	0
46	b	c	d	ab	bc	bc	ab	b	b	bc	b	a	a	bd	a	bc	bcf	ab	bc	dg	bef	bc	abcef	cdeg	cd	bc	ad	ab
	13	13	13	11	13	13	7	13	13	13	13	6	13	13	10	10	12	13	13	13	12	13	13	13	12	13	13	12
47	bd	ac	a–d	abd	c	abc	bc	b	b	cde	b	ad	ab	bde	c	bce	bcdf	a	bceg	dgh	e	ce	bce	beg	cd	ac	a–d	ab
	17	17	17	14	16	17	10	12	17	17	17	2	16	15	13	16	16	17	15	12	16	17	17	17	17	17	14	15

1 = *Acon-1*; 2 = *Acon-2*; 3 = *Acp*; 4 = *Ak-2*; 5 = *Fdp-2*; 6 = *βGal-1*; 7 = *βGal-2*; 8 = *Got-2*; 9 = *αGpd*; 10 = *6-Gpd*; 11 = *Gpi*; 12 = *Hbdh*; 13 = *Idh-1*; 14 = *Idh-2*; 15 = *Ldh*; 16 = *Mdh-1*; 17 = *Me-1*; 18 = *Mpi-2*; 19 = *Odh*; 20 = *PepB-1*; 21 = *PepB-2*; 22 = *PepD-1*; 23 = *PepD-2*; 24 = *Pgm*; 25 = *Pk*; 26 = *Thdh*; 27 = *αAmy*; 28 = *Enol*.